



## D6 – Report on integration of technical system: Serbia



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra



SWISS NATIONAL SCIENCE FOUNDATION

*Deliverable Lead:* FORS  
*Related Work package:* WP1

*Author(s):* Aleksandra Bradić-Martinović (IES)  
Larisa Malić (IES)  
Bojana Tasic (FORS)  
Irena Vipavc Brvar (ADP)  
Maja Dolinar (ADP)

*Dissemination level:* Public (PU)  
*Submission date:* 30<sup>rd</sup> April 2017  
*Project Acronym:* SEEDS  
*Website:* <http://www.seedsproject.ch>  
*Call:* Scientific cooperation between Eastern Europe and Switzerland (SCOPES 2013-2016)  
*Start date of project:* 1<sup>st</sup> May 2015  
*Duration:* 24 months

## Revision History

| Version | Date              | Changes               | Modified by |
|---------|-------------------|-----------------------|-------------|
| 1.0     | February 28, 2017 | Released version      | FORS        |
| 2.0     | April 07, 2017    | Adjustment for Serbia | IES         |
| 2.1     | April 21, 2017    | Revision              | UL- ADP     |
| 3.0     | April 28, 2017    | Final                 | FORS        |

## Acknowledgments

This report has been developed within the “South-Eastern European Data Services” (SEEDS) ([www.seedsproject.ch](http://www.seedsproject.ch)) project. The participant organisations of the SEEDS project are:

| Name  | Short Name | Country       |
|---|------------|---------------|
| Centre for Monitoring and Research, Podgorica   | CeMI       | Montenegro    |
| Centre for Political Courage, Pristina  | CPC        | Kosovo        |
| Institute for Democracy and Mediation, Tirana   | IDM        | Albania       |
| <b>Institute of Economic Sciences, Belgrade</b>   | <b>IES</b> | <b>Serbia</b> |
| Saints Cyril and Methodius University, Institute for Sociological, Political and Juridical Research, Skopje | ISPJR      | Macedonia     |
| University of Zagreb, Faculty of Humanities and Social Sciences, Zagreb                                     | FFZG       | Croatia       |
| Swiss Foundation for Research in Social Sciences, Lausanne  | FORS       | Switzerland   |
| University of Ljubljana, Social Science Data Archive, Ljubljana   | UL-ADP     | Slovenia      |

## Table of Contents

|         |  |    |
|---------|--|----|
| 1       | Introduction.....                                  | 4  |
| 1.1     | OAIS Model.....                                    | 5  |
| 2       | Functional Specifications.....                     | 6  |
| 2.1     | Conceptual Model and Workflow .....                | 6  |
| 2.1.1   | Ingest .....                                       | 6  |
| 2.1.2   | Archival Storage.....                              | 7  |
| 2.1.3   | Data Management.....                               | 9  |
| 2.1.4   | Administration.....                                | 10 |
| 2.1.5   | Preservation Planning.....                         | 11 |
| 2.1.6   | Access .....                                       | 12 |
| 2.2     | Metadata Specifications.....                       | 14 |
| 2.3     | Files and File Formats .....                       | 15 |
| 3       | Technical Specifications.....                      | 18 |
| 3.1     | Tools .....  | 18 |
| 3.1.1   | SEEDSbase .....                                    | 18 |
| 3.1.2   | Dataverse.....                                     | 19 |
| 3.1.3   | Redmine.....                                       | 19 |
| 3.2     | Communication .....                                | 20 |
| 3.2.1   | General Communication .....                        | 20 |
| 3.2.1.1 | Website .....                                      | 21 |
| 3.2.1.2 | Mailing Lists.....                                 | 21 |
| 3.2.1.3 | Direct Contact.....                                | 21 |
| 3.2.2   | Specific Communication .....                       | 21 |
| 3.3     | Technical Infrastructure .....                     | 21 |
| 3.3.1   | Server Architecture (an example) .....             | 21 |
| 3.3.2   | Network and Telecommunications .....               | 23 |
| 3.3.3   | Hardware and Software for production systems ..... | 23 |
| 4       | Conclusions and Future Development .....           | 24 |
|         | Appendix 1: Study description form.....            | 25 |
|         | Appendix 2: Deposit Agreement .....                | 28 |
|         | Appendix 3: End user license.....                  | 30 |

# 1 Introduction

The aim of WP1 of the SEEDS project is to implement the various features of the data service establishment plans. This includes organisational, policy, and technical developments, all geared up toward preparing for “day one” of the new data services in partner countries.

The last activity of WP1 is the integration of the archiving system (chosen in D9 - Report on technical improvements) into the technical infrastructure of the partner institutions. Besides creating a set of policy documents for the data services (see D5 - Policy and procedures document) and new individual websites (see D11), it involves the development of a technical prototype that will allow for the basic archiving functions, following the OAIS model: ingest, preservation, and dissemination. Thus, as a key result of the SEEDS project, institutions in charge for data research archiving (each in their country) have now chosen the tools and have the capacity to take in new social science data, and then to properly document, store, and distribute these data on a national level, all according to international standards. For this purpose, Serbian team established the Serbian Social Science Data Centre (SER.DAC) in 2014 as an organizational unit within the Institute of Economic Sciences, Belgrade (IES). The SER.DAC is output of FP7: SERSCIDA project – Support for Establishment of National/Regional Social Sciences Data Archives (2012-2014). SERSCIDA was designed as a strategic project for supporting the cooperation and exchange of knowledge between the EU countries associated within the Council of European Social Sciences Data Archives (CESSDA) and the Western Balkan Countries in the field of social science data archiving. The project addresses the issues of potential for use of information-communication technologies for the benefit of scientific research and the exchange of knowledge as laid down in the call for proposals. SER.DAC activities are: Preservation and dissemination of quantitative and qualitative data; Provision of access and support for both users and data providers including mediation between the demands of data users and data providers; Permanent monitoring of compliance with the international standards in the field of data management and preservation; Collaboration with the wider scientific community in the field of collecting and distributing data and Open Access Initiative promotion.

This deliverable describes the technical prototype and its related processes. The purpose is to provide the tools and processes that will allow the new data services to begin building their data collections, to structure their data and metadata in ways to allow for discovery and reuse, to store and secure data for the long-term, and to provide the conditions and platforms for data access for their future users. In sum, the prototype supplies a basic archiving infrastructure, with all needed hardware and software.

As has been the case in all previous project outputs, the intention was to maintain as much commonalities as possible across the six new data services, and this is especially true for the established technical platform. Common and compatible tools will allow for future data and information sharing, as well as for synergies across the national services.

## 1.1 OAIS Model

The rapid growth of digital material in both volume and complexity, the rising expectations of archives' users for access services, and the emerging digital preservation strategies, have all contributed to the definition of digital archive functions. The functionalities and procedures of a digital archive have been collected into the OAIS reference model, which became an ISO standard in 2003 (ISO 14721:2003). The OAIS provides both a functional model – the specific tasks performed by the archive, such as storage or access – and a corresponding information model, which includes a model for the creation of metadata to support long-term maintenance and access (see figure 1-1).

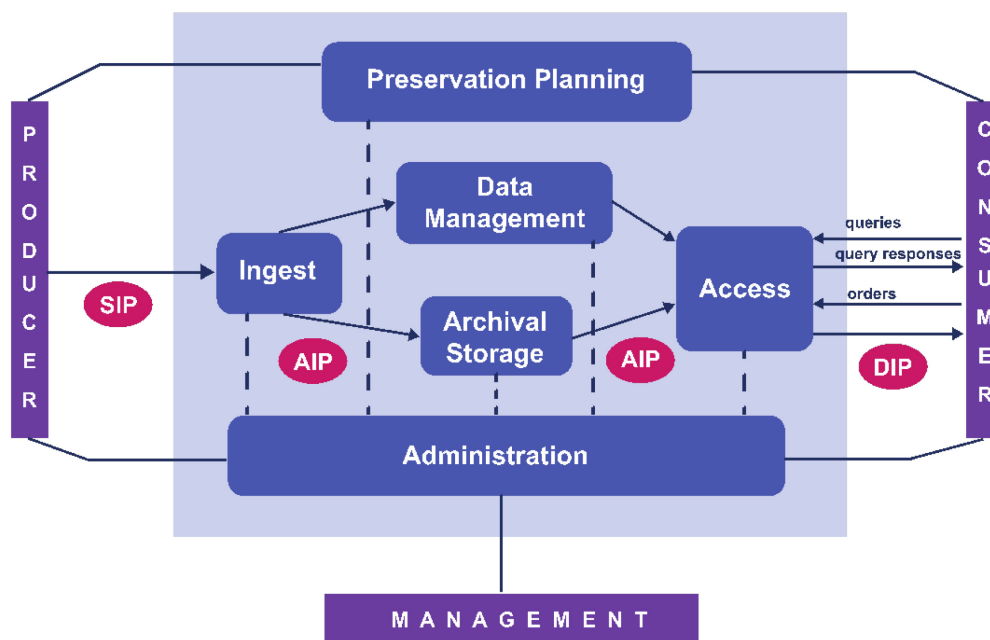


Figure 1-1: OAIS Functional Entities

The OAIS reference model is separated into six functional entities: Ingest, Data Management, Archival Storage, Preservation Planning, Administration, and Access. Outside the OAIS are the Producer (data producers, depositors, researchers), the Consumer (readers, researchers, academics, public, user community), and the Management (data managers, archivists, programmers, database managers, data centre managers). The data within the OAIS are represented as information packages (IPs). Each information package consists of metadata and physical files. There are three types of IPs: submission information package (SIP), archival information package (AIP), and dissemination information package (DIP).

## 2 Functional Specifications

### 2.1 Conceptual Model and Workflow

#### 2.1.1 Ingest

Ingest provides the services and functions to accept SIPs from the Producer and prepare the content for Archival Storage and Data Management within the archive (see figure 2-2).

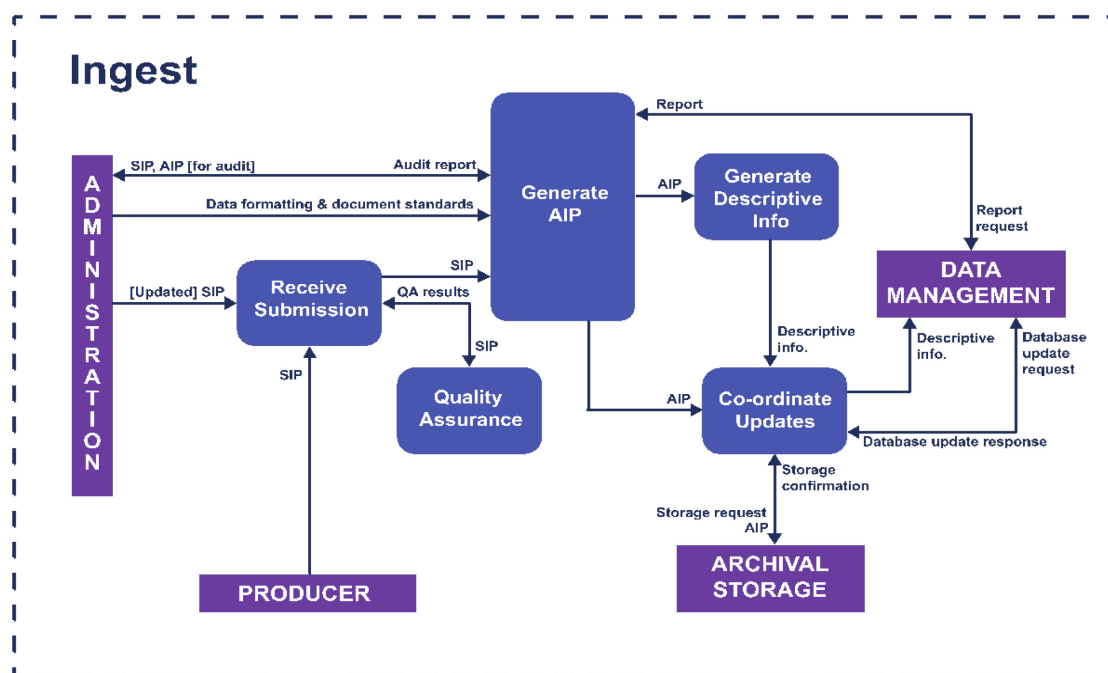


Figure 2-2: Functions of Ingest

For the first contact and negotiation with the Producers about appropriate information for deposit, the SER.DAC is using e-mail communication ([ser.dac@ien.bg.ac.rs](mailto:ser.dac@ien.bg.ac.rs)). The personnel in the Center are trying to obtain sufficient control of the information provided to the level needed in order to ensure Long Term Preservation. Producers have a choice of using encrypted email messages, OwnCloud or personal delivery to the archive (USB stick, CD or similar media devices) for data transfer, depending of the sensitivity of the data. This also refers for all related materials and documentation of the study.

Depositor (producer) has to sign a licence agreement (Deposit Agreement) to establish the terms and conditions of use of his data collection. This is a legal document, which sets out rights and responsibilities of the depositor and the SER.DAC, as the data distributor. The model of Deposit Agreement is enclosed in the Appendix 2.

For collecting metadata, the SER.DAC is using the study description form (Appendix 1) that allow Producers to systematically describe their research studies. In cases, when secondary data sources in

the study are used, only metadata of the study is going to be collected and archived. Reference to original source is going to be made. In addition to the Study description form, the SER.DAC also provides instructions for cleaning and anonymizing data. Besides datasets, the SER.DAC is looking for all relevant documentation, which can provide insight in the data and help other researcher to re-use it (Codebooks, Questionnaires, Structure for interviews or focus groups, Reports, Analysis, Publications). These are published alongside the dataset. SER.DAC also follows these procedures, as a part of the pre-processing phase:

- Review of completeness of the study data and documentation (checked against data deposit agreement);
- Verification of data integrity i.e. has it been fully maintained during the transfer process;
- Checking if confidentiality of data is guaranteed;
- Checking if the scope of documentation is sufficient (e.g. are variables described and can they be easily attributed to the questions in the questionnaire);
- Comparison of data sets with data documentation, in order to ensure that there is a match between variables, variable labels and value labels, and that the number of observations and variables match the stated numbers in the documentation;
- Assignment of an ID-Number to the project within the archiving system (will be replaced by DOI in the future);
- Checking for viruses.
- Checking for data security, from destructive forces and from the unwanted actions of unauthorized users.

As a final step of the Ingest phase, the SER.DAC merges all received data and supported materials and keeps it in their original form as the Submission Information Package (SIP). All data and corresponding documentation for RRPP Data Rescue project are kept on the SEEDSbase platform and our Dataverse platform. As a final choice of our analysis of tools that was made within the deliverable D9 within the SEEDS project, we plan to use only the Dataverse platform in our future activities.

## **2.1.2 Archival Storage**

Archival Storage provides the services and functions for the storage maintenance and retrieval of AIPs (see figure 2-3). Archival Storage functions include receiving AIPs from Ingest and adding them to permanent storage, managing the storage hierarchy, refreshing the media on which archive holdings are stored, migrating files into the archival formats, performing routine and special error checking and providing disaster recovery capabilities.



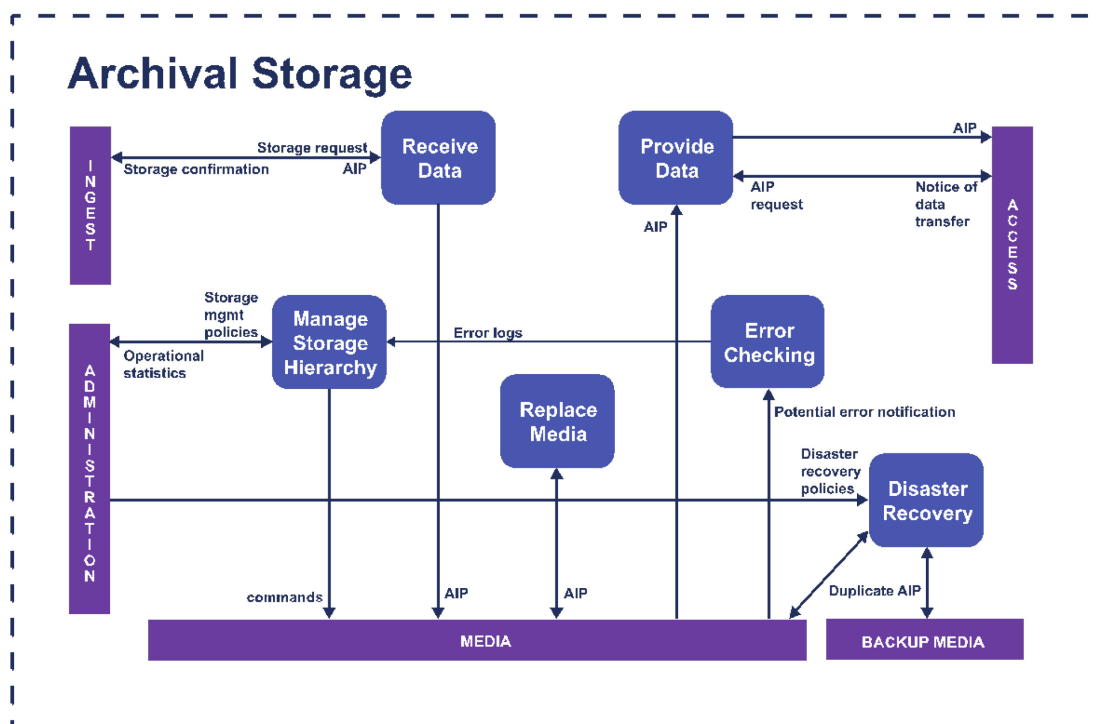


Figure 2-3: Functions of the Archival Storage

The Archival Information Package (AIP) in the SER.DAC is on the same server as the SIP. At the initial phase we used also *OwnCloud* and *Dropbox* storage for internal audit. For the actual archival storage we currently use the SEEDSBase and Dataverse. SER.DAC is using Redmine platform for managing tasks, issues, and requests submitted by a community of users. At the near future we have plans to test Request tracker for this purpose. We have put 10 projects in Redmine so far. There is one main frame: SEEDS-RS, where we put our 10 projects. Every project is formed as *new subproject*. More about this subject can be found in SEEDS D8 deliverable. There is a issue tracking list with trackers:

- Communication
- General
- Materials
- Data
- Study description
- Nesstar
- Before publishing
- Publishing

Recording of preservation tasks in metadata about individual datasets and checksums for authenticity and integrity or log and reports of all refreshment results.

Backup strategies of the SER.DAC archival policy is based on a regular, automatic, daily backup of the server (at 1 a.m).

### 2.1.3 Data Management

Data Management provides the services and functions for populating, maintaining and accessing both metadata, which identify and document repository holdings, and administrative data, used to manage the repository (see figure 2-4).

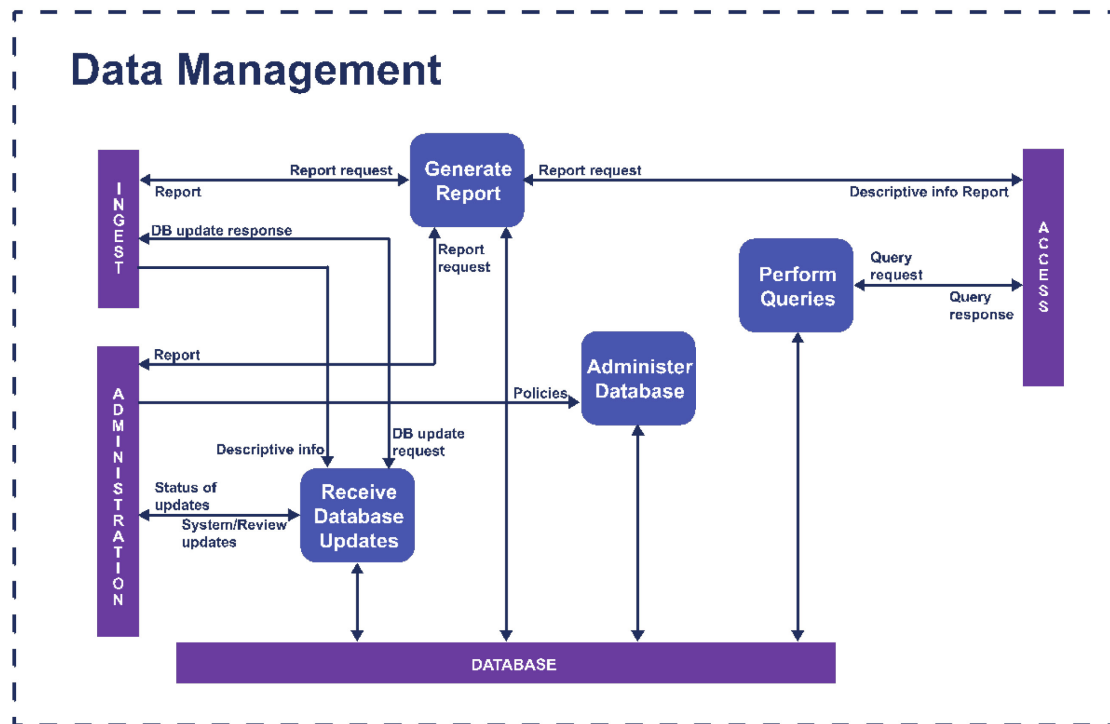


Figure 2-4: Functions of Data Management

The Data Management activities in the SER.DAC will involve the maintenance of all the archived information and files, including the management of metadata, communication with the producers and users, and access statistics. The primary functions of Data Management has been carried out by the archive's staff and include maintaining the databases of metadata for which it is responsible, performing queries on these databases, and generating reports in response to requests from other functional components within the OAIS (Ingest, Administration, Access).

In the SER.DAC the main functionalities of Data Management are integrated into ingest and metadata tools. As we already described in 2.1.2, for this purpose, we used Redmine as a project tracking software tool. We will use it to generate reports based on request from Ingest, Access or Administration and execute any queries or other processes. Our future reports will include summaries of archive holdings by category, or usage statistics for numbers of access of archive holdings.

## 2.1.4 Administration

Administration provides the services and functions for the overall operation of the archive system (see figure 2-5). Administration functions include soliciting and negotiating submission agreements with the Producer, auditing submissions to ensure that they meet archive standards, and maintaining configuration management of system hardware and software. It is also responsible for establishing and maintaining archive standards and policies, providing user support, and activating stored requests.

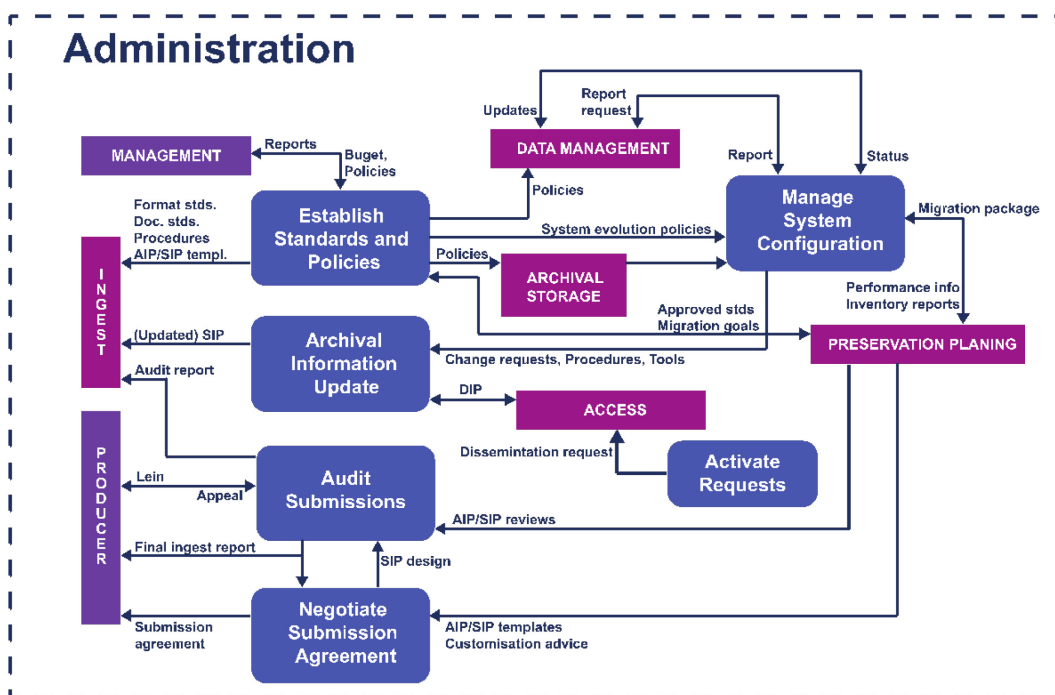


Figure 2-5: Functions of Administration

At the first phase of development, the SER.DAC is using Redmine platform for managing tasks, issues, and requests submitted by a community of depositors. At the near future we have plans to test Request tracker for this purpose.

SER.DAC follows these procedures in pre-release phase:

- Inform the data depositor to check the processed data files and documentation and verify it (quality control) and
- All files created and files changed by archive staff during the ingest procedure (such

as reports and syntax) should be saved and marked as such (preservation metadata).

### 2.1.5 Preservation Planning

Preservation Planning provides the services and functions for monitoring the environment of the archive and making recommendations to ensure that the information stored in the archive remain accessible over a long-term, even if the original computing environment becomes obsolete (see figure 2-6). Preservation planning functions include evaluating the contents of the archive and periodically recommending archival information updates to migrate current archive holdings, developing recommendations for archive standards and policies, and monitoring changes in the technology environment and in the user's service requirements. Preservation Planning also develops detailed migration plans, software prototypes, and test plans to enable implementation of Administration migration goals.

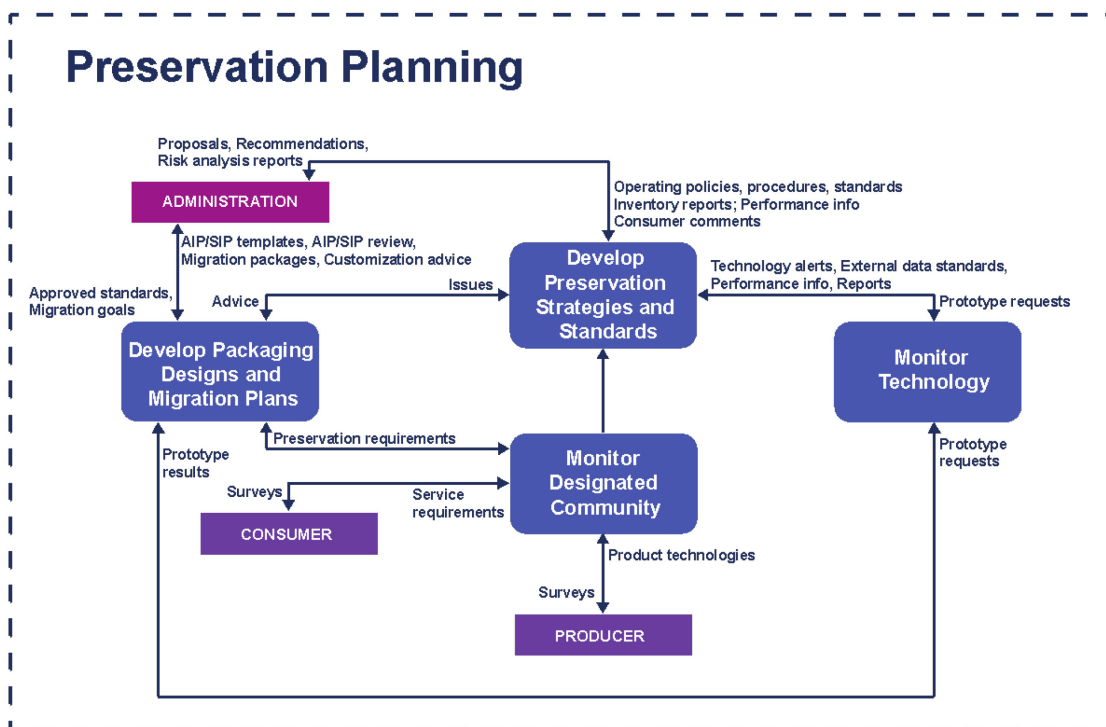


Figure 2-6: Functions of Preservation Planning

SER.DAC did not develop long term plans of its development due to a lack of financing scheme after the SEEDS project, but it has some guidelines. The tasks of Preservation planning will be: (1) development of preservation strategies and standards, (2) development of packaging designs and migration plans, and (3) monitoring of technology (innovations in storage and access technologies) and the designated community (shifts in scope or expectations). In addition to this, MK DASS will monitor the technical fitness of its archive, will do regular risk assessments of the stored digital objects (which includes technology monitoring for the different object types), and plans for

preservation actions.

The SER.DAC is trying to ensure the leading edge of technical advances by taking a strategic approach to long-term digital preservation, and by monitoring hardware and software developments and migrating its collections accordingly. The technical development of storage technology must be under constant observation and review: monitoring technological changes that will affect preservation and migration decisions, files suitable for archiving, the structure of the information packages, and hardware (server, storage media) and software (applications) with support of CESSDA.

Also, the SER.DAC recognises the need for communication with other stakeholders who are active in formulating national preservation policies and programmes. It also acknowledges the need to participate in national and international activities and programmes in the area of digital preservation.

Having in mind that the future of the SER.DAC directly depends on the Ministry's support, we have an intention to make inventory list of the publicly funded scientific projects in the social sciences and to introduce pilot project for archiving and disseminating data. It would help us to build a solid foundation for future work.

### **2.1.6 Access**

Access provides the services and functions that support Consumers in determining the existence, description, location, and availability of information stored in the archive, and in allowing them to request and receive data (see figure 2-7). Access functions include communicating with Consumers to receive requests and applying controls to limit access to specially protected information. This includes coordinating the execution of requests until its successful completion, generating responses, and delivering the responses to Consumers.

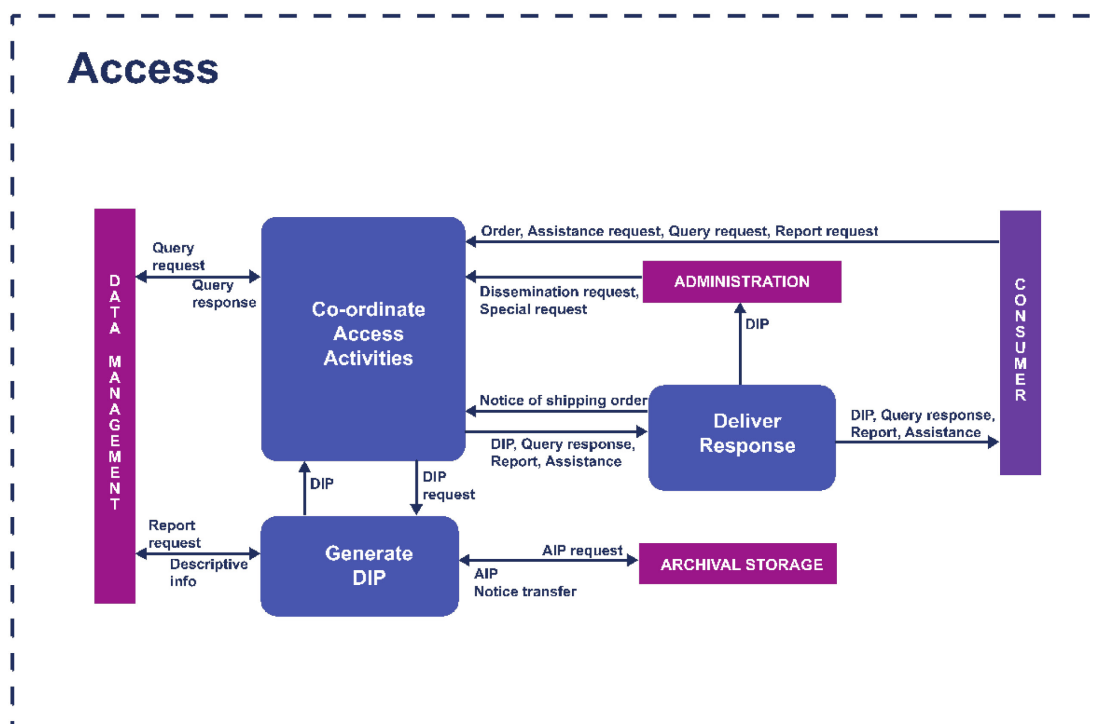


Figure 2-7: Functions of Access

In May, the SER.DAC will have 10 datasets published, and our users will have a chance to access data through the Dataverse portal and through the SEEDSbase portal. The final phase in this process is preparing study collection for download, i.e. **creation of dissemination information packages (DIP)**.

The SER.DAC has two agreements (Deposit and End User) as a formal base for the process of negotiation with third parties. For all researchers, we provided contracts to ensure ours and their security. There is a tension between providing free and open access to data and at the same time protecting confidentiality of research participants and the depositor. However, anonymization strategies on the one side, and specifications in the end user license on the other, are ways to give consideration to both aspects. The access to data at the SER.DAC will be controlled depending on the type of data, the level of sensitivity of the data and the specific conditions determined by the depositor. Metadata and documentations will be publicly available without user registration. Datasets will be publicly available within certain limitations and controlled access:

Access to data is limited to users and individuals affiliated with a research institution, such as faculty members or students (for non-commercial purposes);

Preliminary registration into the system is requested to gain access to data, just for the purpose of evidence (the registration will be done by the users or by archive staff);

All users will have to agree to study-specific usage conditions before they access data (User licence) and

Access to data will be provided by e-mail, direct download link or ftp-server system depending on the sensitivity level of dataset.

There will be three levels of access, depending on the nature of the data:

Data and documents that can be accessed with permission issued by the archive dissemination portals, i.e. registration (Dataverse and SEEDSbase) (Public-Use-File, PUF);

Data and documents that are only available for scientific analysis carried out in academic research and teaching, and may be accessed without impending permission from the data depositor, i.e. free access (download) via ordering system by e-mail or online. However, registration is necessary (Scientific-Use-File, SUF);

Data that are released for academic research and teaching, where the user may gain access to data subject to written approval from the depositor. For this purpose the user must supply the SER.DAC with their details and specify their analytic intention. The SER.DAC will request written permission from the depositor and will only permit access on receipt of this. In any case, the SER.DAC informs the depositor either before or after data distribution to the user depending on their agreement. By default, the user must sign an end-user licence in order to gain access to the material (Secure-Use-File, ScUF).

After the DIP creation, the SER.DAC makes DIP accessible on two platforms: the SEEDSbase for the RRPP data sets and Dataverse for all other data sets. In the processes of obtaining data sets, users have to sign a User Agreement. The User Agreement is enclosed in Appendix 3.

## 2.2 Metadata Specifications

Metadata of a study will be described in Data Documentation Initiative (DDI) metadata specification, version 2.5. Complete documentation is available on the DDI alliance web page<sup>1</sup>.

The DDI is designed to be fully machine-readable and machine processable. It is defined in XML, which facilitates easy Internet access. DDI Controlled Vocabularies<sup>2</sup>, CESSDA topic classification and CESSDA Controlled Vocabularies are planned to be used. The use of a predefined topic classification will make possible future inclusion in the CESSDA data catalogue<sup>3</sup> easier, since these are the topics that enable browsing in the catalogue.

The fields for ingest in the archive ingest tool are made using the CESSDA-recommended fields<sup>4</sup>, relevant for study description.

---

<sup>1</sup> <http://www.ddialliance.org/Specification/DDI-Codebook/2.1>

<sup>2</sup> <http://www.ddialliance.org/controlled-vocabularies>

<sup>3</sup> <http://www.cessda.net/catalogue/>

<sup>4</sup>

<https://cessda.net/content/download/709/6350/file/CESSDA%20mandatory%20and%20recommended%20metadata%20fields.pdf>

## 2.3. Files and File Formats

There are several reasons why a data archive should be concerned with file formats: they exist in big numbers, are relevant during the whole workflow of the OAIS reference model, and are largely proprietary. File formats are subject to rapid obsolescence if they are not evaluated according to crucial criteria, such as open standards, ubiquity, interoperability, and metadata support. Therefore, file formats that are well-documented, non-proprietary and usable on different hardware and software platforms are much less at risk of not being usable anymore in the future. In addition, their frequency of migration and their costs of preservation are lower.

File formats are an important issue during the entire workflow of the archive (see chapter 2.1). In the functional entity Preservation Planning, the composition and attributes of the information package are defined. This includes the selection of file formats for the SIP, the AIP and the DIP. The decisions of the archive on which file formats are acceptable as archival and distribution formats are linked to the significant properties of the files (what aspects of the digital material we want to preserve). That is why it is important that file formats are controlled and validated, according with the available specific tools, already in the Ingest phase.

There are a number of tools on the market for migrating a file format into a more reliable and sustainable file format:

- *Native Java Image library* for most image formats;
- *Imagemagick* for most image formats, esp. Raster;
- *FFMPEG* for various AV formats;
- *readpst* for email;
- *Ghostscript* for PDF;
- *Libre Office* for Office Open XML, and word processor files – also shifts various office formats to PDF and PDF/A;
- *Inkscape* for Vector images.

When selecting target formats, the following criteria should be considered:

- Ubiquity;
- Support;
- Disclosure;
- Documentation quality;
- Stability;
- Ease of identification and validation;
- Intellectual Property Rights;
- Metadata Support;
- Complexity;
- Interoperability;
- Viability;
- Re-usability.



The selected file formats represent a summary of different recommendations from CESSDA partners and internationally recognised institutions: <sup>5</sup>

In addition, file format registries are a way of helping to identify file formats and looking up format specifications.

SER.DAC follows international best practices in its choice of preservation formats. Preservation procedures will consist of:

- Reformatting data sets from proprietary software to software-independent formats and creating an Archival Information Package (AIP);
- Defining, timing, testing and implementing migration pathways;
- When new formats are created from data files either through migration into new file formats or through creating new file formats for dissemination, the old files are retained alongside.

SER.DAC uses the following formats for the AIP:

- Tabular data: Microsoft Excel File Format (XLS) (.xls), ASCII, Comma Separated Values (CSV) (.csv; .txt);
- Text: Adobe Portable Document Format (PDF/A) (.pdf), XML (.xml), Rich Text Format (RTF) (.rtf), Standard Generalised Markup Language (SGML) (.sgml);
- Audio: MP3 (.mp3);
- Images: TIFF (.tif), MPEG-2 (.mpg2).

Formats for the DIP are:

- Tabular data: SPSS portable format (.por), SPSS (.sav), Stata (.dta);
- Text: Adobe Portable Document Format (PDF) (.pdf), Rich Text Format (RTF) (.rtf);
- Audio: MP3 (.mp3);
- Images: JPEG (.jpg).

The SER.DAC also includes in its Policies and procedures preferred formats for the receipt of digital materials:

- Tabular data : SPSS portable format (.por), SPSS (.sav), Stata (.dta), Excel or other spreadsheet format files, which can be converted to tab- or comma-delimited text;
- Text: Adobe Portable Document Format (PDF/A, PDF) (.pdf), plain text data, ASCII (.txt);
- Audio: Waveform Audio Format (WAV) (.wav) from Microsoft, Audio Interchange File Format (AIFF) (.aif) from Apple, FLAC (.flac);
- Images: TIFF (.tif) ideally version 6 uncompressed, JPEG (.jpeg, .jpg) only when created

---

<sup>5</sup> The formats highlighted in bold are preferred over the others of the same category.

FORS: Qualitative Data Archiving at FORS – Policy and Procedures:

[http://www2.unil.ch/daris/IMG/pdf/Donnees\\_qualitatives\\_archivees\\_chez\\_FORs\\_-\\_Politique\\_et\\_Procedures.pdf](http://www2.unil.ch/daris/IMG/pdf/Donnees_qualitatives_archivees_chez_FORs_-_Politique_et_Procedures.pdf),

UK Data Archive: Formats table: <http://www.data-archive.ac.uk/create-manage/format/formats-table>, UK Data Archive: Assessment of UKDA and TNA Compliance with OAIS and METS Standards, p. 89

<http://www.dptp.org/wp-content/uploads/2010/08/UKDAp90.pdf>.

in this format, Adobe Portable Document Format (PDF/A, PDF) (.pdf), RAW image format (.raw), Photoshop files (.psd);

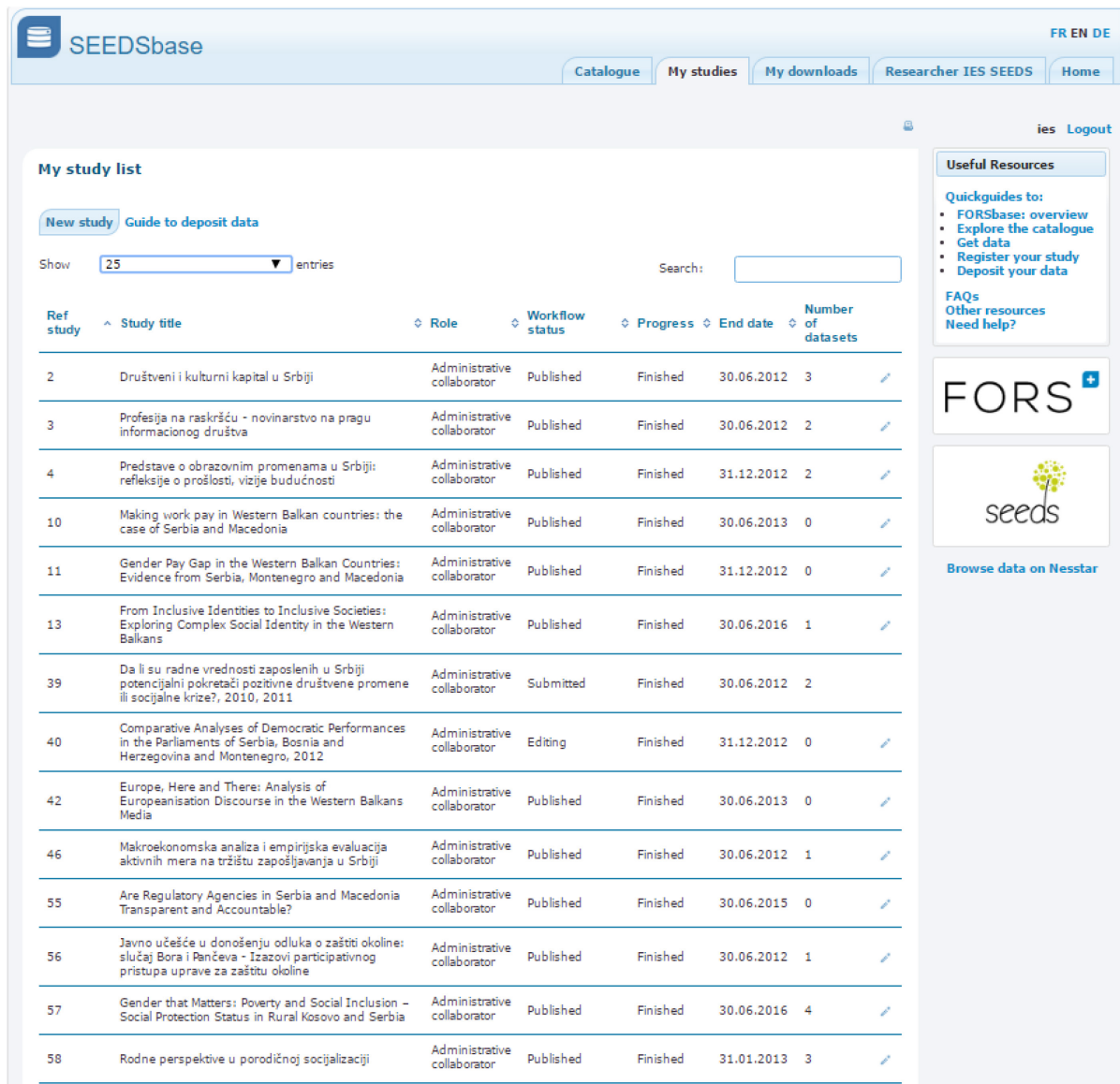
- Video: MPEG-4 (.mpg4), motion JPEG 2000 (.mj2);
- Compressed files: are accepted as long as they can be uncompressed by using open and freely available software, such as 7-Zip or Winzip.

## 3 Technical Specifications

### 3.1 Tools

#### 3.1.1 SEEDSbase

During the RRPP Data Rescue project, as a main platform for managing SIP and DIP, the SER.DAC used SEEDSbase to archive 23 studies with quantitative and qualitative datasets.



The screenshot displays the SEEDSbase web interface. At the top, there is a navigation bar with the SEEDSbase logo and language options (FR EN DE). Below the navigation bar, there are tabs for 'Catalogue', 'My studies', 'My downloads', 'Researcher IES SEEDS', and 'Home'. The main content area is titled 'My study list' and includes a search bar and a dropdown menu for the number of entries (set to 25). A table lists 23 studies with columns for 'Ref study', 'Study title', 'Role', 'Workflow status', 'Progress', 'End date', and 'Number of datasets'. The table contains the following data:

| Ref study | Study title   | Role                        | Workflow status | Progress | End date   | Number of datasets |
|-----------|---|-----------------------------|-----------------|----------|------------|--------------------|
| 2         | Društveni i kulturni kapital u Srbiji   | Administrative collaborator | Published       | Finished | 30.06.2012 | 3                  |
| 3         | Profesija na raskršću - novinarstvo na pragu informacionog društva  | Administrative collaborator | Published       | Finished | 30.06.2012 | 2                  |
| 4         | Predstave o obrazovnim promenama u Srbiji: refleksije o prošlosti, vizije budućnosti  | Administrative collaborator | Published       | Finished | 31.12.2012 | 2                  |
| 10        | Making work pay in Western Balkan countries: the case of Serbia and Macedonia   | Administrative collaborator | Published       | Finished | 30.06.2013 | 0                  |
| 11        | Gender Pay Gap in the Western Balkan Countries: Evidence from Serbia, Montenegro and Macedonia  | Administrative collaborator | Published       | Finished | 31.12.2012 | 0                  |
| 13        | From Inclusive Identities to Inclusive Societies: Exploring Complex Social Identity in the Western Balkans                            | Administrative collaborator | Published       | Finished | 30.06.2016 | 1                  |
| 39        | Da li su radne vrednosti zaposlenih u Srbiji potencijalni pokretači pozitivne društvene promene ili socijalne krize?, 2010, 2011      | Administrative collaborator | Submitted       | Finished | 30.06.2012 | 2                  |
| 40        | Comparative Analyses of Democratic Performances in the Parliaments of Serbia, Bosnia and Herzegovina and Montenegro, 2012             | Administrative collaborator | Editing         | Finished | 31.12.2012 | 0                  |
| 42        | Europe, Here and There: Analysis of Europeanisation Discourse in the Western Balkans Media  | Administrative collaborator | Published       | Finished | 30.06.2013 | 0                  |
| 46        | Makroekonomska analiza i empirijska evaluacija aktivnih mera na tržištu zapošljavanja u Srbiji  | Administrative collaborator | Published       | Finished | 30.06.2012 | 1                  |
| 55        | Are Regulatory Agencies in Serbia and Macedonia Transparent and Accountable?  | Administrative collaborator | Published       | Finished | 30.06.2015 | 0                  |
| 56        | Javno učešće u donošenju odluka o zaštiti okoline: slučaj Bora i Pančeva - Izazovi participativnog pristupa uprave za zaštitu okoline | Administrative collaborator | Published       | Finished | 30.06.2012 | 1                  |
| 57        | Gender that Matters: Poverty and Social Inclusion - Social Protection Status in Rural Kosovo and Serbia                               | Administrative collaborator | Published       | Finished | 30.06.2016 | 4                  |
| 58        | Rodne perspektive u porodičnoj socijalizaciji   | Administrative collaborator | Published       | Finished | 31.01.2013 | 3                  |

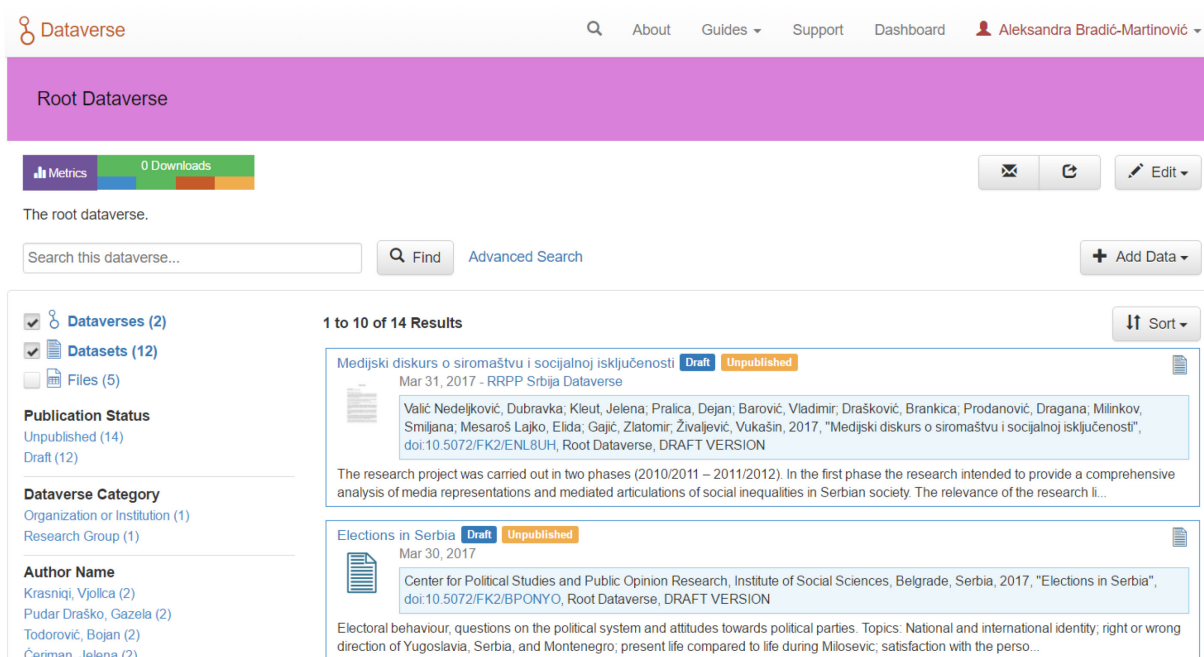
On the right side of the interface, there is a 'Useful Resources' section with links to 'Quickguides to: FORSbase: overview, Explore the catalogue, Get data, Register your study, Deposit your data', 'FAQs', and 'Other resources'. Below this, there are logos for 'FORS' and 'seeds', and a link to 'Browse data on Nesstar'.

Based on the Study description document (Annex 3), received from PIs, we filled all required fields in the SEEDSbase and uploaded all data files (in SPSS format for quantitative data and pdf format for qualitative data). We also included all received documentation in pdf (questionnaires, codebooks, additional explanation, etc.) and all available publications (reports, monographs, articles, etc.). After the quality check done by FORS and the ADP team, the data had been published.

We faced two main issues with the SEEDSbase. Firstly, the Serbian team did not have an administrator account for this platform, so we did not have full access to all phases in our curation process, i.e. we were depending on the FORS team. In this case, we had to deal with the SEEDSbase with depositor account. Secondly, the SEEDSbase platform is not an open source software at the moment, and was developed to fit FORS work process best. After our software evaluation (SEEDS Deliverable D9 - Report on technical improvements) we choose Dataverse as our main archiving tool.

### 3.1.2 Dataverse

According to the SEEDS project's plan, the Serbian team had to archive 10 datasets. We used Dataverse to accomplish this task (<http://dataverse-serbia.ien.bg.ac.rs/dataverse.xhtml>).



The screenshot shows the Dataverse web interface. At the top, there is a navigation bar with the Dataverse logo, a search bar, and links for About, Guides, Support, and Dashboard. The user's name, Aleksandra Bradić-Martinović, is displayed in the top right corner. Below the navigation bar, there is a purple header with the text "Root Dataverse". A metrics bar shows "0 Downloads". There are buttons for "Metrics", "0 Downloads", "Email", "Share", and "Edit". Below this, there is a search bar with the text "Search this dataverse..." and buttons for "Find" and "Advanced Search". There is also a button for "Add Data". The main content area shows a list of search results. On the left, there are filters for "Dataverses (2)", "Datasets (12)", and "Files (5)". Below these are filters for "Publication Status" (Unpublished (14), Draft (12)), "Dataverse Category" (Organization or Institution (1), Research Group (1)), and "Author Name" (Krasniqi, Vjolica (2), Pudar Draško, Gazela (2), Todorović, Bojan (2), Čeriman, Jelena (2)). The search results are displayed in a table with columns for "1 to 10 of 14 Results" and "Sort". The first result is "Medijski diskurs o siromaštvu i socijalnoj isključenosti" (Draft, Unpublished) by Valić Nedeljko, Dubravka; Kleut, Jelena; Pralica, Dejan; Barović, Vladimir; Drašković, Brankica; Prodanović, Dragana; Milinkov, Smljana; Mesaroš Lajko, Elida; Gajić, Zlatimir; Žvaljević, Vukašin, 2017, "Medijski diskurs o siromaštvu i socijalnoj isključenosti", doi:10.5072/FK2/ENL8UH, Root Dataverse, DRAFT VERSION. The second result is "Elections in Serbia" (Draft, Unpublished) by Center for Political Studies and Public Opinion Research, Institute of Social Sciences, Belgrade, Serbia, 2017, "Elections in Serbia", doi:10.5072/FK2/BPONYO, Root Dataverse, DRAFT VERSION.

Although we managed to publish all studies, the selected platform is still under adjusting. The main area for improvement we would like to see in the future would be multiple language option and additional metadata fields, which will be in accordance with the DDI metadata scheme and CESSDA recommended metadata fields and vocabularies.

### 3.1.3 Redmine

The SER.DAC is using the Redmine platform for managing tasks, issues, and requests submitted by our community of users. In the near future we have plans to test the Request tracker for these purposes.

Home My page Projects Help

SEEDS-RS » Search: » Rodne perspektive u porodičnoj socijalizaciji, 2012

## Rodne perspektive u porodičnoj soci...

+ Overview Activity **Issues** Gantt Calendar News Documents Wiki Files Settings

Issues + New issue

Filters

Status open Add filter

Tracker is Communication

Options

Apply

| #                            | Tracker       | Status | Priority | Subject  | Assignee       | Updated             |
|------------------------------|---------------|--------|----------|--|----------------|---------------------|
| <input type="checkbox"/> 136 | Communication | New    | Normal   | 1.2. Discuss access conditions, data protection and deposit contract | Jelena Banović | 03/20/2017 03:42 PM |
| <input type="checkbox"/> 135 | Communication | New    | Normal   | 1.1. Receiving files   | Jelena Banović | 03/09/2017 03:58 PM |
| <input type="checkbox"/> 134 | Communication | New    | Normal   | 1. Communication with the data provider                              | Jelena Banović | 03/09/2017 03:56 PM |

(1-3/3)

Also available in: [Atom](#) | [CSV](#) | [PDF](#)

You can read more about it at [D8](#).

## 3.2 Communication

### 3.2.1 General Communication

According to the OAIS model there are several different possibilities for how the data archive can communicate with the actors, that is the data producers and consumers. More precisely, it is the functional entities Preservation Planning and Administration that are responsible for the communication task. They include for instance the development of preservation strategies and standards of monitoring the community and technology in order to meet the needs of the producers and consumers (see figure 3-1).

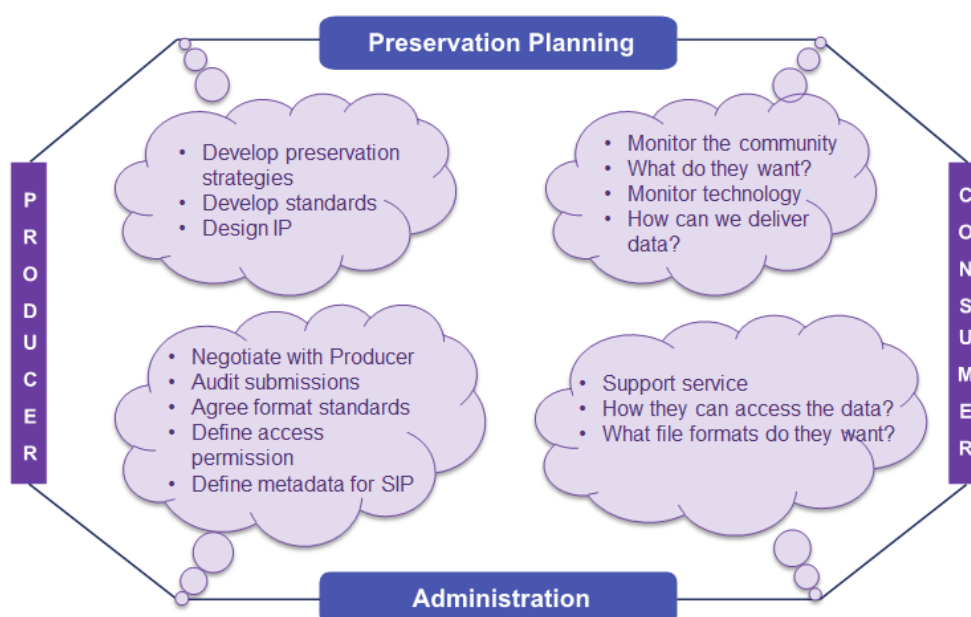


Figure 3-1: Communication

### 3.2.1.1 Website

The most common and wide-reaching channel to communicate with the community is by means of an institutional website. It is the showcase for interested consumers and producers of data to learn about how data can be obtained and submitted. It is the platform where the policy and procedures, reports and publications, guidelines for data preparation, description of data protection, and other training materials are made available.

The website of SER.DAC can be visited on: <https://rs.seedsproject.ffzg.hr/>

### 3.2.1.2 Mailing Lists

A mailing list is established by the SER.DAC in order to inform Producers and Consumers about the latest news and upcoming events, such as training and workshops. We are using [ser.dac@ien.bg.ac.rs](mailto:ser.dac@ien.bg.ac.rs) email address.

### 3.2.1.3 Direct Contact

A third way of communication is direct contact of the Producers and the Consumers or potential users of the data service through sporadic interaction on an as-needed basis (e.g., for workshops, seminars, and conferences).

## 3.2.2 Specific Communication

All the specific communications with the users during user registration and Ingest will be recorded and maintained via Redmine, as we already described.

## 3.3 Technical Infrastructure

### 3.3.1 Server Architecture (an example)

For the implementation of the SEEDS project we need one Virtual server. The Virtual server is used for the hosting of each national web portal. Currently, a single WordPress application with 6 website instances (one for each partner) is installed for the national web portals (see [D11](#)).

Here is the detailed Virtual server configuration:

**Configuration:** 2 vCPU, 2GB RAM, 10GB HDD

**OS:** Debian GNU Linux 8.2 (Jessie)

**HA:** Ganneti cluster<sup>6</sup>

Virtual server should have redundant IT infrastructure, monitoring, and backup. In addition to local (on-site) file and database backup, there should be a daily automatic offsite backup solution as well. The local servers should be used for backup purposes.

The usage of virtual machines is valuable for prototype implementation and testing, but for the production system, the newly established archives should have more granular distribution of services. Sensitivity of data in the various components of the OAIS, requires us to think about different security levels of data and preservation requirements. To achieve this goal, the future architecture will be installed separately on different virtual machines, based on different platform deployment stacks:

Website,

Dataverse,

Redmine.

Each of these components have different deployment requirements (database, web server, runtime language stack), so it makes sense to separate components on different VMs to enable easy maintenance (migration when changing components, deploying different components for new archives in the future, firmware upgrades).

Looking at the current state of development and support probability of chosen software of the established data archives, it seems that a future change in the components will be probable. This is one of the reasons why the easy maintainability of the system is important. The staff of the archive needs to be capable of testing other available software tools (in a state accessible to them), preferably under Free/Libre/Open Source licences, by using the process described in deliverable D9- Report on technical improvements.

Since each application is installed on a separate virtual machine (and each might have its own set of issues/bugs), security issues are addressed for each virtual machine individually. This means for example that in case of security problems on the web portal, there will be no effect on the security of the archival copy of the data or any other component of the archival infrastructure.

All virtual machines should have two copies stored on different physical machines locally. Machines should be located in different buildings to ensure continuous operation in case of environmental problems in one of the buildings (fire, flooding etc.).

During the process of developing an OAIS based data archive, two distinct types of data required for keeping in the archive were identified - SIP and AIP, which require long-term preservation together with an audit log. This also requires the ability to check whether data is correctly stored on the media

---

<sup>6</sup> <http://www.ganeti.org/>

that requires checksums on the level of the file system (scrubbing). For this requirement, ZFS<sup>7</sup> storage and snapshots using LVM could be implemented to provide a long-term archival copy of current prototype on different locations (e.g. in faculty building), which should be updated daily (from computing centre location). This would enable disaster recovery in case of one location failure. It is also possible to have multiple remote copies, if needed.

The management of applications and data could be done using Ganeti<sup>8</sup>, an open source cloud solution that enables high availability for virtual machines and provides data storage requirements outlined above.

### **3.3.2 Network and Telecommunications**

The network infrastructure and telecommunications are accessed using the host organisations' systems: Serbian National Academic Network (AMRES) and local infrastructure within the Institute of Economic Sciences in Belgrade.

### **3.3.3 Hardware and Software for production systems**

Based on best practices and international standards for social science data archives, the SER.DAC has determined the hardware and software it will use.

Workstation computers that will be used by future archive staff for Data Management should include the following software: office tools; conversion tools; software for statistical analysis (STATA, R, SPSS); tools for preparing metadata description of a study, etc.

If we decide to use a proprietary product, we will have to buy a licence or use the existing licences of our hosting institution, if available.

---

<sup>7</sup> <http://bit.ly/dc14-zfs>

<sup>8</sup> <http://bit.ly/dc14-ganeti>



## 4 Conclusions and Future Development

In conclusion, the prototype described in this report provides the technical basis for all key archiving functions, following the OAIS model. The new data services will be able to integrate and manage new datasets, safely store and protect data, as well as disseminate data and documentation to users. Their technical systems will function according to international norms and best practices, even if some of the archiving workflow will need to be handled manually.

It should be noted, however, that while the prototype will enable certain basic services, it will not be as comprehensive or as flexible as the one used by mature social science data archives. Future work should expand the technical development to accommodate for a greater volume and variety of data, to automate more everyday practices, and to enhance communication potential and exchange with data producers and users. This work will continue for many years, and will build on experience, further training, and funding. Like any others, these new data services will have to adapt technically to the ever-changing research and policy environments.

## Appendix 1: Study description form

**Study title:** \_\_\_\_\_

**Study language:** \_\_\_\_\_

**Institutions:**

Please enter here the full name of the institution(s) that carry out the research and provide the address and possibly its website. Also enter the name of mandating institutions if applicable.

**Keywords:**

(Maximum of 8 research domains or key words)

**Disciplines:**

(Mark all that apply)

Administration and management sciences

Anthropology

Communication and media sciences

Criminology

Demography

Economics

Educational sciences

Environmental sciences

Ethnology

History

Human and cultural geography

Legal science

Political sciences

Psycho- and sociolinguistics

Psychology

Social and economic history

Social and preventive medicine

Social policy

Social psychiatry

Social psychology

Social work

Sociology

Urbanism and spatial planning

Other disciplines (specify) \_\_\_\_\_

**Period:**

(Please enter the period covered by the project, e.g. 1939-1945. This is not necessarily the same as the length of the research project!)

**Geographical space:**

(If applicable, eg. Macedonia, Serbia, Albania)

**Abstract:**

**Results**

(Please enter the main or intermediate results. This part can also be updated during the term of the project)

**Method description:**

(Describe briefly the methods used and the steps followed)

**Methods (instruments):** **COPY THIS PAGE AS MANY TIMES AS INSTRUMENTS ARE USED**

(If data are gathered or analysed, indicate for each collection carried out the various methodological aspects in reference to the list below. If it describes qualitative research, please replace the items in the list with the appropriate information.)

**Method instrument (select/mark one, then describe below):**

- Analysis of content
- Analysis of content (standardized)
- Analysis of content (open)
- Analysis of documents
- Analysis of documents (standardized)
- Analysis of documents (open)
- Laboratory instrument
- Psychological test
- Observation
- Observation (participant)
- Focus group
- Qualitative interview
- Standardized interviews
- Standardized interviews (face to face)
- Standardized interviews (by telephone)
- Standardized interviews (by mail)
- Secondary analysis
- Secondary analysis (individual data)
- Secondary analysis (aggregate data)
- Other (specify)

\_\_\_\_\_

- Observation (without participation)

**Population, unit of observation:**

**Participant selection or sampling method:** (census, random sampling, other sampling method (specify)):

**Other significant comments:**

**Publications:**

(Please note the publications (monographs, journal articles, reports, etc.) that directly resulted from the research project, indicating the complete reference (author(s), title, series or journal, editor, place and date of the publication, URL, DOI, etc.))

**Unpublished documents:**

(Please enter any unpublished documents (presentations, unpublished reports, working papers, etc.) that are directly related to the research project. Please indicate the complete reference.)

**Secondary analyses:**

(Is the publication based on a secondary analysis of the dataset.)

**Start - End date:**

(Please indicate the planned or actual dates of the start and the end of the research project. For managing updates, an indication of the end date is essential, even if it is only temporary. For periodic surveys that are continuing, the end date may remain free.)

## Appendix 2: Deposit Agreement

### 1. The institution depositing the data, "INSTITUTION"

Represented by "RESEARCHER"

- a. maintains the copyright on the data in point 2 as well as on the corresponding documentation,
- b. has collected and treated the data in conformity with existing laws on data protection,
- c. confirms that the data have been anonymised, and
- d. deposits them at SER.DAC, Institute of Economic Sciences, Belgrade, Zmaj Jovina 12.

### 2. The data to be archived

Data and documentation from the project entitled "STUDY", under the direction of "RESEARCHER (PI)".

### 3. The goal of the archiving of the data

The deposit of the data and documentation at SER.DAC has the aim of:

- a. their archiving at SER.DAC,
- b. their reuse by third parties for the purpose of analysis.

### 4. Responsibilities of SER.DAC

SER.DAC agrees to:

- a. archive the data and the documentation in accordance with its possibilities and technical standards,
- b. make the data and documentation available to third parties for analytic purposes,
- c. inform data users of their contractual and legal obligations, which are notably:
  - to cite completely the data source, the institution, and the researchers that produced the data,
  - to use the data and instruments only for their own analytical purposes. In particular, it is not permissible to transmit the data to third parties, whether in original or modified form, whether for reward or free of charge.

SER.DAC is however not responsible for possible violations of these terms by data users.

### 5. Rights of SER.DAC

- a. SER.DAC has the right to publish information on the data and the corresponding research project in order to inform about the archived data and to promote their reuse,
- b. SER.DAC has the right to provide to third parties, for analytic purposes, copies of the archived data and the corresponding documentation,
- c. SER.DAC has the right to use the archived data for analytic and to publish results.

## 6. Rights of the institutions that deposits the data

- a. The institution has the right to be informed about any reuse of the data by third parties.

## 7. Special conditions for depositing the data

---

Depositor (Institution/Organisation):

The SER.DAC

Place: "PLACE "

Date: "DATE"

Belgrade, the "DATE"

## Appendix 3: End user license

### I agree to the following:

1. to use the data cited above and related instruments only for scientific research and teaching within an academic framework and for no other purpose and to use the data for the purpose that I have outlined in my description in the present contract ;
2. to use the data with respect to national law and standard norms of data protection;
3. not to undertake to attribute data to specific individuals; and to communicate obtained results only in a way that does not identify individual cases (i.e., persons, households, or institutions);
4. to respect above all individual confidentiality, and not to divulge any information that might compromise this confidentiality;
5. to use the data in a conscientious and informed manner, notably by consulting the related documentation, and to respect scientific ethical rules of conduct;
6. to cite used data and documents in conformity with scientific standards (using the citation above);
7. to inform SER.DAC of all publications based on these data;
8. to not transmit these data to third parties, whether in original or modified form, whether for reward or free of charge, except to individuals who have also signed this contract;
9. to store the data in a way that no third party can gain access to it;
10. to destroy the data after analyses (as outlined in the description below) are completed, and to confirm this to SER.DAC before publication of results.

For what purpose will the data be used? :

Brief description of the expected research:

Current status:

Expected publication:

Duration of use:

### Additional conditions for highly sensitive data:

The cited data have not been anonymised or contain information easily permitting identification of individual cases. Therefore, use of the data requires special care with respect to data protection. In addition to the points above, **I agree to the following:**

1. To respect above all individual confidentiality, and not to divulge any information that might compromise this confidentiality.
2. To transmit these conditions to third parties, if need be, who are eligible to work with these data within the framework of the project for which these data were obtained.

The data user:

“ADDRESS”

Serbian Data Center in Social Sciences

Place:

Belgrade, the “DATE”:

Signature: